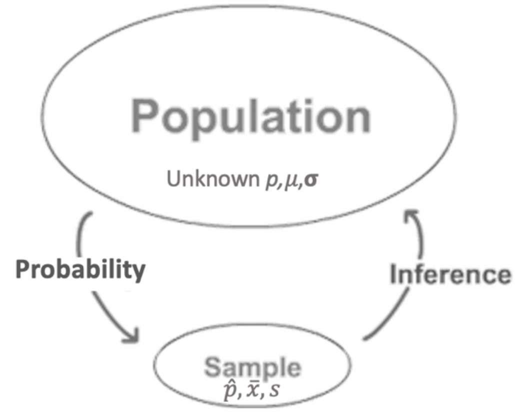
11.5 Confidence Intervals – Overview

Inference

- **Inferential statistics** involves using descriptive statistics to estimate population parameters.
- The GOAL is to learn about a POPULATION.

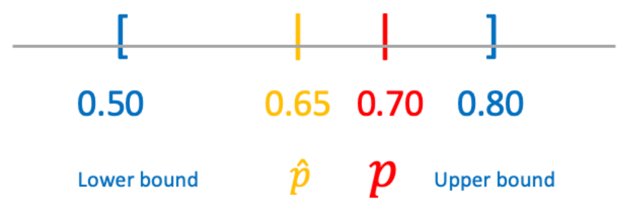


Why Confidence Intervals?

- GOAL: Estimate Parameters

- Point Estimates
  - Using a statistic to estimate a parameter (this means we use $\hat{p}$ or $\bar{x}$ to estimate p or μ, respectively)
  - It is a single number that is our best guess (estimate).
  - Very unlikely that statistics equal the true parameter values they are estimating (remember each sample is different; sampling variability).
  - Therefore, in order for the estimate to be useful, we must describe how close it is likely to be.

- Interval Estimates
  - Give a range for what we think the population parameter is.
  - Takes into account sampling variability.



How to Build a Confidence Interval

Point Estimates

- **Example 1**: Calculate the point estimates for the following scenarios.

  - The campus bookstore is determining if they need to increase their marketing budget. In order to check this, they took a random sample of 137 students in which 81 students said they buy their books at the campus bookstore.

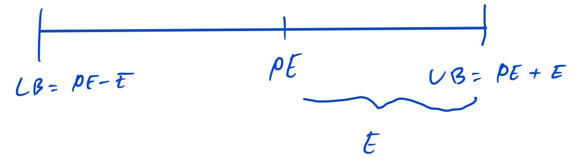
$$\hat{p} = \frac{x}{n} = \frac{81}{137} \approx 0.591$$

  - Scientists discovered a new mountain range under the sea. From a random sample of 13 peaks, there they recorded the following heights.

Height
7700
11500
16000
5800
9000
5500
12400
22100
14200
19300
8300
9100
6100

$$\bar{x} = \frac{\sum x}{n} = \frac{7,700 + 11,500 + \cdots + 6,100}{13} = 11,308$$

Margin of Error



C.I. = Point Estimate ± Margin of Error

- Point Estimate is your best guess; at the center of the interval.
  - Then we extend our guess in both directions in order to provide a wider range of plausible values.
  - This distance is called the **Margin of Error**.

- Margin of Error (E) is what makes our estimates intervals rather than just single points!

- **Example 1 Continued**: Calculate the confidence interval based on the following scenarios.

  - Using $\hat{p} = 0.591$ as the point estimate for the proportion of students who purchase books at the campus bookstore, suppose we know the margin of error is 0.023.

    CI = [LB, UB]   PE ± E = 0.591 ± 0.023 = [0.568, 0.614]

  - Using $\bar{x} = 11{,}308$ ft as the point estimate for the mean height of the new underwater mountain range, suppose we know the margin of error 4,500 ft.

    CI = [LB, UB] = PE ± E   11,308 ± 4,500 = [6808, 15808]

Surveys

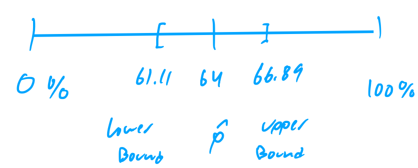
- A common use of confidence intervals is for estimating proportions based on surveys.
- In this context, we can approximate the margin of error based on the sample size.

- **Rule of thumb for margin of error in a survey**
  - With 95% confidence, the margin of error, E, is approximately

    $\boxed{\frac{1}{\sqrt{n}} \cdot 100\%}$ for a sample of size n

- **Example 2**: A NatGeo Poll interviewed 1200 hiking enthusiasts and asked "Are you more afraid of spiders or snakes???" Out of the 1200 people, 768 responded "Ewww, snakes….". Calculate the corresponding margin of error.

  $\frac{1}{\sqrt{1200}} \cdot 100\% \approx 2.89\%$



How to Interpret a Confidence Interval

Interpretation

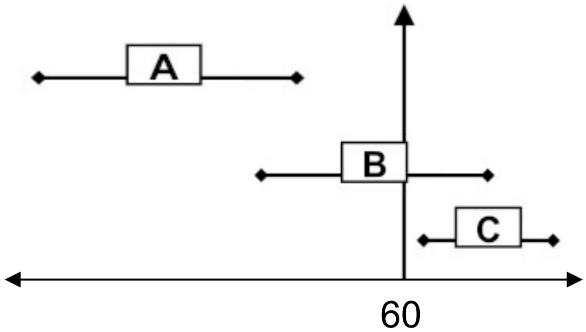
- A General Structure
  - I am C% confident that the true/population parameter + context is between (lower bound) and (upper bound).

- **Example:** Trying to estimate the proportion of all Muncie residents who enjoy running
  → 95% CI = (0.05, 0.25)

  - We are 95% confident that the true (population) proportion of all Muncie residents who enjoy running is between 0.05 and 0.25.

Comparing Confidence Intervals

- When comparing confidence intervals to a particular value, or other intervals, we need to look at the ENTIRE interval to see if it is COMPLETELY below or above our comparison.



<u>Comparisons</u>

A << 60

B ?? 60

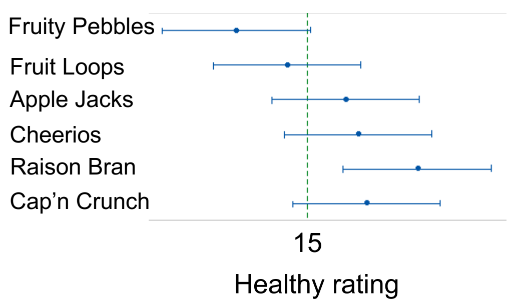
C >> 60

A ?? B

A << C

B ?? C

- **Example**: Determine the total number of significant comparisons that we can conclude from the given confidence intervals.
    - In other words, how many confidence intervals are below / above 15 and/or below / above each other?



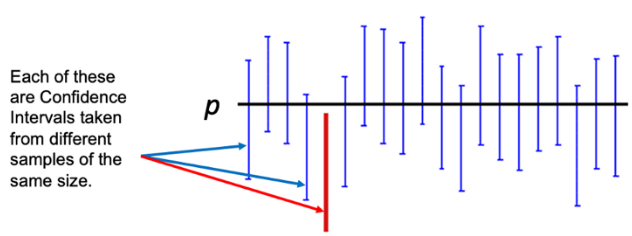
<u>Comparisons</u>

Raison Bran >> 15

Raison Bran >> Fruity Pebbles

Level of Confidence

- The level of confidence, denoted by *c*, is the percentage of all possible samples of a given size that will produce interval estimates that contain the actual parameter.



A 95% confidence interval indicates that 19 out of 20 samples (95%) from the same population will produce confidence intervals that contain the population parameter.

- **Very Important!**
    - The confidence level is NOT the probability the parameter is in the interval.
    - It refers to the long run capture rate (i.e. over many, many intervals constructed in the same way).
    - Either the interval contains the parameter, or it does not.

- How to choose the level of confidence?
  - Tradeoff between Precision and Confidence
    - If we want to be more confident, we need to cover more values!
    - That of course decreases the precision…
    - So there is a pro / con to being super confident.